Analyse statistique des séries chronologiques

M1. MAS, M. BOUTAHAR, M. CARENZI

Contenu

Une série chronologique Z_t peut être modélisée en considérant les trois composantes suivantes:

- La tendance, notée m_t, elle traduit le fait que la valeur de la série croit ou décroit dans le temps t
- La saisonalité notée $\mathbf{S}_t$, elle permet de décrire le phénomène cyclique de la série
- Terme d'erreur, notée B_t, il modélise la partie aléatoire de la série.

Selon la nature de la composante cyclique, on peut utiliser l'un des deux modèles suivants:

- Modèle additif: $Z_t = m_t + \mathbf{S}_t + B_t$
- Modèle multiplicatif: $Z_t = m_t * \mathbf{S}_t * B_t$

En général si l'amplitude du cycle varie dans le temps on utilise le modèle multiplicatif, phénomène que l'on peut observer sur le chronogramme de la série, cependant il existe plusieurs tests statistiques qui permettent de décider si le modèle et additif ou multiplicatif.

Contenu

Nous allons présenter le test de Buys-Ballot qui est simple à programmer:
Soient n la taille de la série Z_t, et r la saison, posons N=n/r; (on suppose que n est un multiple de r). On définit alors la matrice

$$Y[i,j] = Z[i+(j-1)*r], 1 \leq i \leq r, 1 \leq j \leq N.$$

et les deux vecteurs des moyennes et des écart-types de l'année j suivants

$$Moyenne[j] = \frac{1}{r}\sum_{i=1}^{r} Y[i,j],$$

$$Sigma[j] = \sqrt{\frac{1}{r}\sum_{i=1}^{r}(Y[i,j] - Moyenne[j])^2}, 1 \leq j \leq N.$$

On effectue la régression des écart-types sur les moyennes

$$Sigma[j] = b_1 + b_2 Moyenne[j] + \epsilon_j.$$

Si le coefficient b_2 est significatif (p-value inférieur à 0.05) alors on décide que le modèle est multiplicatif sinon il est additif.

Exemple 1. On considère la série des ventes suivante:

Trimestre / Année	2015	2016	2017
T_1	1248	891	1138
T_2	1392	1065	1456
T_3	1057	1118	1224
T_4	3159	2934	3090

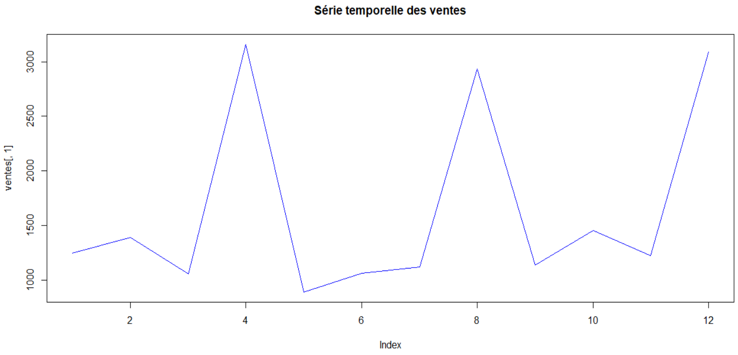


Figure 1: Chronogramme des ventes

L'application du test de Buys-Ballot nous donne une p-value=0.7864, par conséquent le modèle est additif.

Une fois le test de Buys-Ballot effectué, si le modèle est additif on prends $X_t = Z_t$ alors que s'il est multiplicatif on prendra $X_t = \log(Z_t)$, avec cette transformation logarithmique le modèle multiplicatif deviendra un modèle additif. Dans la suite nous allons donner la démarche de modélisation par un modèle additif.

Contenu

On se limite dans ce cours à identifier une tendance linéaire de la forme $m_t = b_1 + b_2 t$, pour ceci on effectue la régression suivante:

$$X_t = b_1 + b_2 t + B_t, 1 \leq t \leq N,$$

Si la p-value associée à b_2 est inférieur à 0.05 alors la tendance est significative, c'est à dire que la chronique croît (si $b_2 > 0$) et décroît (si $b_2 < 0$) avec le temps.

- Si la tendance est significative alors on prendra $m_t = \hat{b}_1 + \hat{b}_2 t$, où $\hat{b}_1$ et $\hat{b}_2$ sont les estimateurs de b_1 et b_2 respectivement.
- Si la tendance n'est pas significative alors on prendra $m_t = \overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$.

Exemple 1 (suite). Si on considère la série des ventes, l'application de la régression nous donne une p-value =0.4213 pour b_2 et donc la tendance n'est pas significative par conséquent $m_t = \overline{X} = 1647.6667$

Contenu

Dans ce cours nous allons présenter deux méthodes à savoir:

- La méthode des moyennes saisonnières
- La méthode des moyennes mobiles

Contenu

Pour identifier la composante saisonnière on considère la série sans tendance $W_t = X_t - m_t$.
La composante saisonnière est supposée de la forme:

$$\mathbf{S}_t = S_i \text{ si } (\text{t mod r}) = \text{i} \ , 1 \leq i \leq r, 1 \leq t \leq n \qquad (1)$$

(t mod r) est le reste de la division de t par r.
Pour que le modéle soit identifiable on impose la condition suivante sur les coefficients saisonniers $S_1, ..., S_r$

$$S_1 + S_2 + ... + S_r = 0. \qquad (2)$$

On définit la matrice

$$Y[i,j] = W[i + (j-1) * r], 1 \leq i \leq r, 1 \leq j \leq N.$$

le vecteur des coefficients saisonniers est donné par

$$S'[i] = \frac{1}{N} \sum_{j=1}^{N} Y[i,j];$$

Par la suite on normalise les coefficients saisonniers pour satisfaire la condition (2):

$$S = S' - \overline{S'}, \text{ avec } \overline{S'} = \frac{1}{r} \sum_{i=1}^{r} S'[i].$$

Exemple 1 (suite). On considère la série des ventes, on obtient alors les coefficients saisonniers suivants:

S_1	-555.3333
S_2	-343.3333
S_3	-514.6667
S_4	1413.3333

Interprétation :

-La moyenne des ventes sur la période analysée 2015-2017 est de $m_t = 1647.6667$, à chaque trimestre les ventes vont augmenter ou diminuer de la valeur de S_i selon le signe de celui-ci, par exemple pour le trimestre 4, les ventes vont augmenter de 1413.3333 par rapport à la moyenne alors que pour le trimestre 1, les ventes vont chuter de -555.3333.

-La chronique désaisonnalisée est définie par:

$X_t^s = X_t - \mathbf{S}_t$, avec $\mathbf{S}_t = S_i$ si (t mod r) =i, $1 \leq i \leq r$.

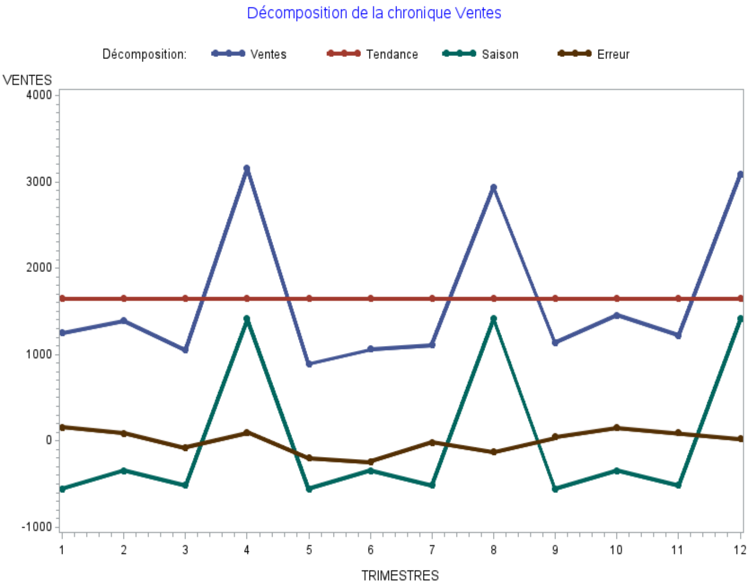
-La chronique corrigée de la tendance est donnnée par:

$X_t^d = X_t - m_t$.

-La série résiduelle (partie aléatoire ou stochastique) est donnée par :

$R_t = X_t - m_t - \mathbf{S}_t$.

La figure 2 montre les trois composantes $m_t, \mathbf{S}_t, R_t$ ainsi que la série brute X_t de la série des ventes.


Décomposition de la chronique Ventes
Décomposition:
Ventes
Tendance
Saison
Erreur
VENTES
4000
3000
2000
1000
0
-1000
1
2
3
4
5
6
7
8
9
10
11
12
TRIMESTRES

Contenu

Pour identifier la composante saisonnière on considère la série sans tendance $W_t = X_t - m_t$.

La taille de l'échantillon doit être multiple de r, et celui-ci doit être pair.

Posons $m = r/2, N = n/r$; on définit le vecteur des moyennes mobiles par

$$\begin{aligned} Z[t] &= \frac{1}{r}\left\{\frac{1}{2}W[t-m] + W[t-m+1]\right. \\ &+ \left. ... + W[t+m-1] + \frac{1}{2}W[t+m]\right\}; \\ & (m+1) \leq t \leq (n-m); \end{aligned}$$

le vecteur A de dimension n

$$A[i] = \begin{cases} W[i] - Z[i] & \text{si } (m+1) \leq t \leq (n-m) \\ 0 & \text{sinon} \end{cases}$$

et la matrice

$$Y[i,j] = A[i + (j-1) * r], 1 \leq i \leq r, 1 \leq j \leq N.$$

Le vecteur des coefficients saisonniers (non normalisés) est donné par

$$S'[i] = \frac{1}{N-1} \sum_{j=1}^{N} Y[i,j], 1 \leq i \leq r;$$

Le vecteur des coefficients saisonniers normalisés est alors donné par

$$S = S' - \overline{S'}, \text{ avec } \overline{S'} = \frac{1}{r}\sum_{i=1}^{r} S'[i].$$

Exemple 1 (suite). Si on considère la série des ventes alors les coefficients saisonniers sont donnés par

S_1	-592.7188
S_2	-352.8438
S_3	-508.1563
S_4	1453.7188